2К

УДК 681.3:621.921.34:001.895:621.9.02

В.Н. Кулаковский, А.А. Лебедева, К.З. Гордашник, Е.М. Чистяков, И.В. Скворцов
Институт сверхтвердых материалов им. В.Н. Бакуля НАН Украины, г. Киев, Украина
vnk@ism.kiev.ua

Онтологический подход к построению базы знаний «Сверхтвердые материалы»

Предложен подход к построению многоуровневой онтологии для системы знаний предметной области «Сверхтвердые материалы (СТМ)». Рассматривается сложно структурированная предметная область сверхтвердых материалов, определяются компоненты базы знаний СТМ и процесс ее разработки на основе метаонтологии «Сверхтвердые материалы (СТМ)». Авторами разработан модульный подход к созданию модели онтологии и показано его применение на примере небольшого фрагмента иерархической модели предметной области «Поликристаллы синтетических алмазов». Предлагается создание онтологии в виде совокупности модулей, где каждый модуль описывает терминологию некоторого раздела предметной области. В работе дается краткий анализ предметной области «Сверхтвердые материалы», ее деление на разделы, описываются онтологии и модели двух разделов – «Природные СТМ» и «Синтетические СТМ».

Введение

В настоящее время в сети Интернет представлен большой объем знаний и информационных ресурсов по сверхтвердым материалам (СТМ). Однако эти ресурсы достаточно хаотично распределены по каталогам и электронным архивам или размещены на отдельных сайтах, что значительно затрудняет их поиск и использование. Кроме того, неструктурированные форматы представления информации (предназначенные только для чтения), не позволяют осуществлять программный поиск новой информации. Как показывает анализ публикаций [1-4], одним из перспективных подходов к повышению эффективности как поиска, так и анализа информации, является подход, основанный на построении онтологий предметной области. Представляя собой не что иное, как практическую реализацию единой модели знаний предприятий и организаций, онтология способствует интеграции разнородных информационных ресурсов в рамках системы на концептуальном уровне, обеспечивая единый подход к описанию их семантики.

1. Основные понятия

Под онтологиями будем понимать совместно используемые, формальные классификации предметной области (ПрО). Формальная онтология – универсальная модель знания – представляет собой систему понятий (объектов), свойств этих объектов, отношений между этими объектами и утверждений, построенных из этих объектов, их свойств и отношений между ними. База знаний (БЗ) онтологической системы состоит из предметной онтологии, онтологии задач и метаонтологии. Объектами предметной онтологии являются объекты предметной области, объектами онтологии задач являются задачи предметной области, а метаонтология формально описывает структуру предметной онтологии и онтологии задач.

Онтологии являются моделями данных, обладающих двумя специфическими особенностями:

1. Онтологии строятся на основе совместного понимания предметной области в рамках сообщества. Это понимание представляется соглашением экспертов по поводу понятий и отношений, которые имеются в предметной области.

2. Онтологии используют способ представления, который может обрабатываться компьютерными программами (т.е. записываются с использованием формальных языков), что дает возможность компьютерам работать с онтологиями. К таким действиям относятся передача онтологий между компьютерами, хранение онтологий, проверка согласованности онтологий, выполнение логических выводов на онтологиях и с помощью онтологий.

Модель онтологии – это система логических соотношений, каждое из которых имеет смысловое содержание, с которым согласно некоторое сообщество экспертов (специалистов предметной области), а вся система есть явное представление знаний предметной области.

Описание ПрО «СТМ» включает совокупность терминов и отношений, семантически значимых для данной предметной области, а также правил, согласно которым можно строить утверждения об элементах ПрО. Модель онтологии предметной области является основой многоуровневой программной системы, позволяющей оперативно осуществлять интеграцию, выборку и управление знаниями в области современных сверхтвердых материалов.

Основные концепции исследований онтологии ПрО.

1. При построении концептуальной модели используются предметные знания в виде набора понятий и связывающих их отношений. Каждое понятие имеет имя и атрибуты, каждый атрибут имеет значение.

2. Модель построена с использованием следующих средств: словаря терминов, используемых при описании характеристик объектов и процессов, точных и однозначных определений всех терминов этого словаря и классификации логических взаимосвязей между этими терминами.

Базисом проектирования концептуальной модели онтологий является ПрО «СТМ», обладающая хорошо отработанным и выверенным понятийным аппаратом. ПрО имеет корректно построенную иерархию предметно-ориентированных определений и является основой концептуального и информационного моделирования.

Концептуальной моделью ПрО будем называть модель, разработанную без ориентации на используемые в дальнейшем программные и технические средства.

Концептуальная модель предметной области указывает на то, какая информация будет содержаться и обрабатываться в проектируемой системе, не касаясь вопросов, как это будет реализовано. Структура данных, описывающая предметную область на смысловом уровне, является проблемно-ориентированной и системно-независимой, то есть независимой от конкретной СУБД, операционной системы и аппаратного обеспечения.

Определим основные понятия, используемые для построения концептуальной модели:

1. Понятие объекта.

Объект – это такое абстрактное множество предметов, в котором все предметы в этом множестве – экземпляры – имеют одни и те же характеристики. Все экземпляры подчинены и согласованы с одним и тем же набором правил и линий поведения. Каждый объект в информационной модели должен быть обеспечен

уникальным именем и идентификатором. Объекты могут абстрагироваться в класс объектов для представления объектов специализированного назначения, имеющих общие атрибуты.

2. Понятие атрибута.

Все предметы в реальном мире имеют характеристики. Каждая отдельная характеристика, которая является общей для всех возможных экземпляров объекта, абстрагируется в отдельный атрибут.

Атрибут – это абстракция одной характеристики, которой обладают все абстрагированные как объект сущности. Каждый атрибут обеспечивается именем, уникальным в пределах объекта. Множество атрибутов может объединяться в группу атрибутов и иметь идентификатор группы атрибутов. Множество идентификаторов групп может быть объединено в класс и иметь идентификатор класса.

3. Представление.

Объект, вместе со своими атрибутами, в информационной модели представляется в графическом или текстовом виде.

4. Понятие связи.

Связь – это абстракция набора отношений, которые систематически возникают между различными видами предметов в реальном мире. Реальные предметы должны быть сами абстрагированы как объекты. Каждая связь должна иметь уникальный идентификатор.

Для фиксации значимых отношений между терминами выделялись основные связи между ними, которые графически можно отобразить с помощью диаграмм отношений. Такие связи в дальнейшем могут послужить основой для интеграции различных онтологий.

2. Структура БЗ «Сверхтвердые материалы»

Современная реализация СУБД института – набор приложений, выполняющихся на сервере баз данных, доступ к которым с рабочих мест пользователей осуществляется через локальную сеть с использованием терминального доступа средствами Windows 2003 Terminal Server.

В качестве примера построения нынешних БД рассмотрим структуру одной из баз данных – каталог «Материалы и инструменты из СТМ». Он состоит из трёх физически не связанных реляционных БД основных СТМ, каждая из которых обладает сложной структурой, а именно – совокупностью из четырёх таблиц (литература, разработчики, производители, свойства), связанных отношением один ко многим с главной таблицей (название и аналоги). Помимо этого в базу входит еще две таблицы – назначение и способ получения (рис. 1).

В базе данных не отражена взаимосвязь между свойствами материалов, способом получения и назначением. Но подобные связи существуют, и их определение и поиск закономерностей в этих связях важны как при создании новых материалов, так и при использовании старых. Связь между разработчиками, производителями и литературой, возможно, не так очевидна, но, тем не менее, важна при патентном поиске, поиске направлений будущих исследований, а также для понимания новых тенденций исследований в этой области.

Решает эту проблему изменение структуры базы, но, поскольку БД с отражёнными в ней причинно-следственными связями и закономерностями взаимодействий уже является базой знаний, то для ее построения следует использовать

соответствующие современные инструменты и стандарты. Это позволяет обеспечить совместимость с широким спектром имеющегося на рынке ПО для работы с БЗ, а также возможность сотрудничества между научными организациями.

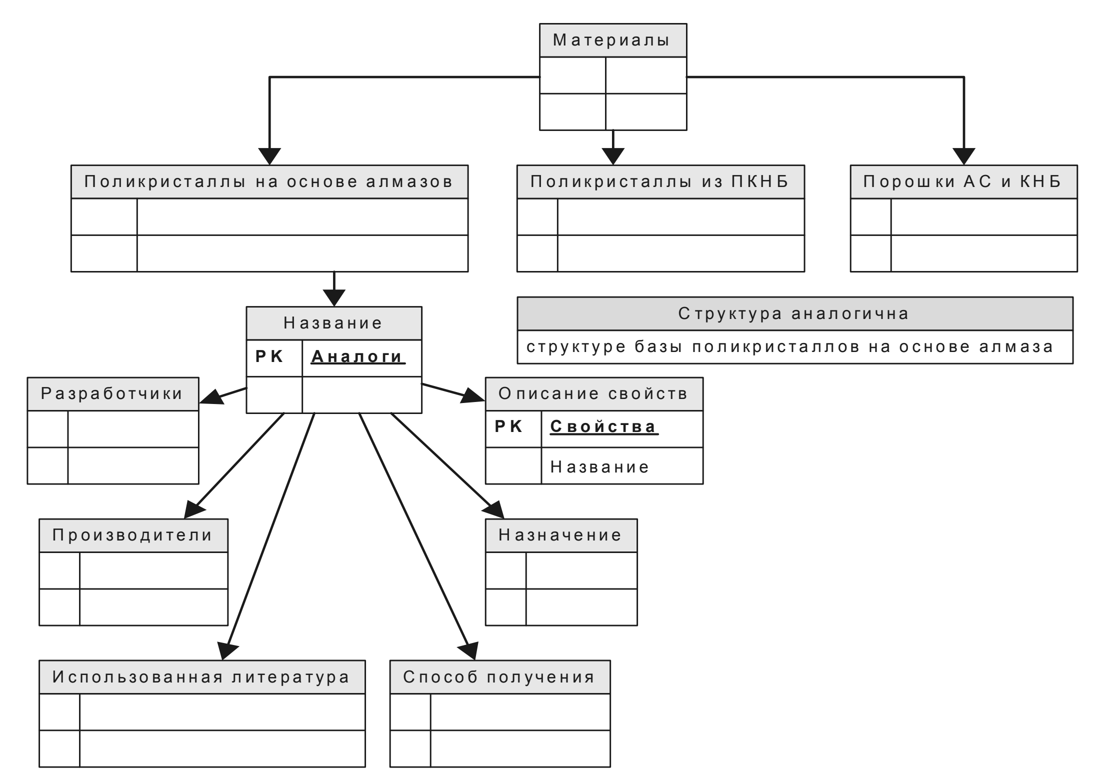


Рисунок 1 – Структура каталога «Материалы и инструменты из СТМ»

Для организации удалённого доступа к базам применяются web-ориентированные технологии.

Произведено обновление программного обеспечения системы управления базами данных и знаний института, смены формата хранения информации и смены идеологии построения базы. Таким образом, обеспечены:

– взаимосвязь между базами данных и знаний,
– разделение баз и программных интерфейсов к ним,
– унификация форматов данных,
– возможность удалённого доступа,
– лёгкость переноса данных на другие аппаратно-программные платформы.

По нашему мнению, всем этим параметрам отвечает построение базы знаний института с использованием онтологического подхода. Построив онтологию предметной области и заполнив ее конкретными экземплярами классов, мы получаем полную взаимосвязанность между чисто справочной информацией (свойства материала), технической информацией (особенности создания материала, например – режим спекания, катализатор, исходные вещества), научной информацией (физические и химические процессы, протекающие во время создания материала), экономической (стоимость конечного материала, трудо- и энергозатраты) и патентно-публикационной (кто, когда получил вышеупомянутые данные, в каких источниках это указано).

База знаний «Сверхтвердые материалы» состоит из двух подсистем первого уровня: «Природные СТМ» и «Синтетические СТМ» [5], [6]. Подсистема второго уровня «Алмазные СТМ» включает две подсистемы третьего уровня: «Виды СТМ» и

«Области применения», которые содержат наиболее характерные сведения о видах природных алмазных СТМ и областях их применения. В подсистему второго уровня «СТМ на основе синтетических алмазов» входят три подсистемы третьего уровня, каждая из которых содержит кроме подсистем «Виды СТМ» и «Области применения» третью подсистему – «Способы получения».

На рис. 2 показан фрагмент общей логической структуры ПрО «Сверхтвердые материалы».

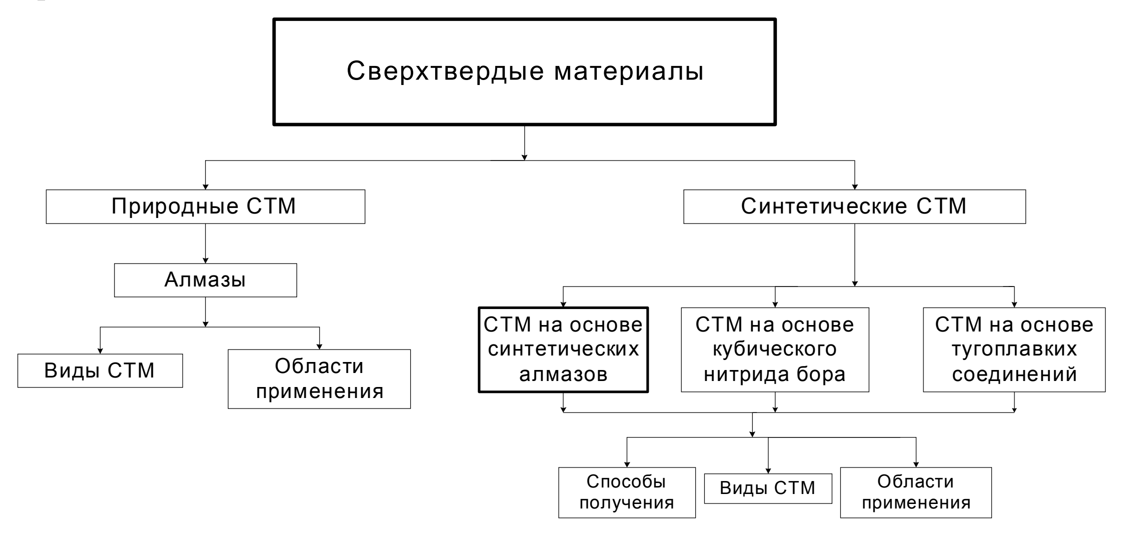


Рисунок 2 – Фрагмент общей логической структуры базы знаний «Сверхтвердые материалы»

Система формирования модулей последующих подчиненных уровней БЗ «Сверхтвердые материалы» показана на рис. 3.

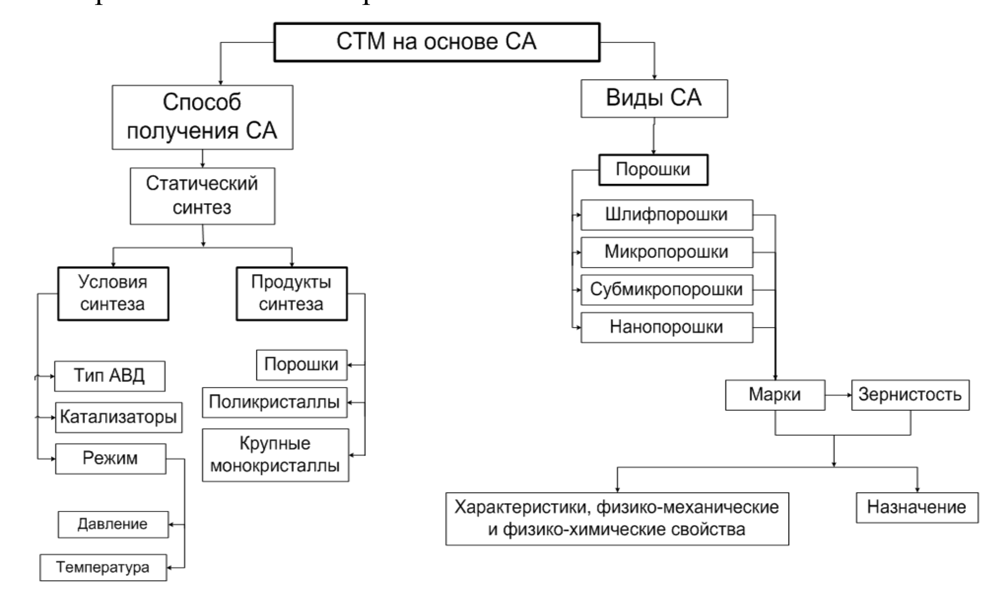


Рисунок 3 – Фрагмент логической структуры уровня «Сверхтвердые материалы на основе синтетических алмазов»

В процессе разработки данной модели использовались базы данных прикладных задач рассмотренной ПрО.

3. Построение онтологии для анализа данных о сверхтвердых материалах

Процесс построения онтологии состоял из нескольких этапов.

Первым шагом в создании онтологии была разработка терминологического тезауруса (около 2000 понятий) в области СТМ. Анализ нашими экспертами полученного тезауруса показал, что для решения прикладных задач (например, многофакторного анализа по прогнозированию свойств СТМ) описания только свойств и структуры сверхтвердых материалов недостаточно. Такое описание является адекватным для классификации материала на основании его физико-механических свойств. Однако для прогнозирования способов и режимов его получения, а также возможностей дальнейшего использования необходимы дополнительные данные. Поэтому предметная область была дополнена подобластями «Способы и режимы спекания порошков» и «Области применения», которые, по сути, являются отдельными предметными областями.

После этого на естественном языке создавали список точных определений терминов. Затем на основе таксономических отношений строили деревья классификации понятий.

Модель онтологий СТМ содержит как многоуровневые материаловедческие понятия, так и проблемно-ориентированные знания, соответствующие научным направлениям института сверхтвердых материалов (ИСМ). Система онтологий знаний и понятий в области СТМ является информационной основой для построения базы знаний «СТМ».

В процессе разработки онтологии выделяются и формально описываются классы понятий, связанные в иерархию с помощью отношения наследования. Различные свойства каждого понятия описываются с помощью атрибутов понятий и ограничений, наложенных на область их значений. Механизм наследования определен таким образом, что наследующему понятию от родительского понятия передаются не только все атрибуты, но и отношения.

Для разработки общей логической структуры системы онтологий «Сверхтвердые материалы» использовали модульный принцип (рис. 2).

Описанный подход построения онтологии апробирован на примере небольшого фрагмента иерархической модели ПрО «Поликристаллы синтетических алмазов». Были использованы описанные выше основные концепции (положения) предметной области и определены отношения между ними. Процесс построения онтологий состоит из создания следующих блоков:

1. Классов и их свойств (classes, properties).
2. Свойств каждой концепции, описывающих различные функциональные возможности и атрибуты концепции (слоты (slots).
3. Ограничений по слотам (также известных как аспекты/грани (slot facets), иногда называемые ограничения ролей).

Модули организованы в иерархию, т.е. модуль может быть модулем верхнего уровня или подчиненным модулем, ассоциированным с отдельным объектом.

Экспериментальная модель онтологии «Поликристаллы синтетических алмазов» была создана в среде Protege-2000 [7], которая является интегрированным инструментальным средством и успешно используется разработчиками систем и экспертами предметных областей знаний.

Для создания онтологии была иерархически упорядочена и структурирована вся понятийная область «Поликристаллы синтетических алмазов», включая способы получения материалов, используемые типоразмеры и разновидности исходных порошков синтетических алмазов (по зернистости – от наноразмерных до крупнозернистых, по составу и свойствам – от алмазных порошков без специальных добавок к порошкам с добавками и наполнителями). Кроме того, определены режимы получения ПКА с использованием техники высоких давлений и температур, а также соответствующих методов спекания порошков СА, а также прямого превращения графитовых материалов в заготовки ПКА.

На рис. 4 показан фрагмент общей логической структуры ПрО «Поликристаллы синтетических алмазов».

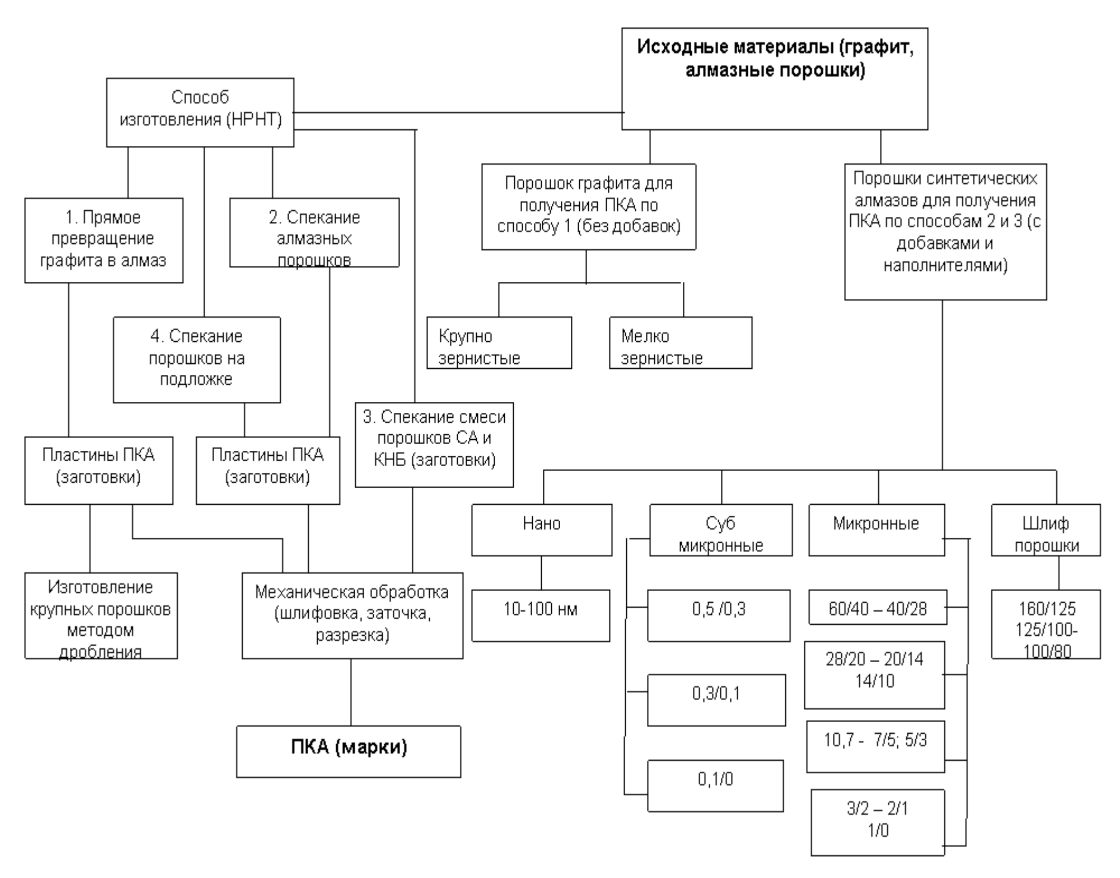


Рисунок 4 – Структура онтологии «Поликристаллы синтетических алмазов»

В онтологию «Поликристаллы синтетических алмазов» включены такие понятия, как способ изготовления, режим спекания, зернистость исходных порошков, физико-механические свойства ПКА, марки ПКА.

Таким образом, для выбранной нами базовой проблемно-ориентированной области «Поликристаллы синтетических алмазов» построена структурная схема онтологии (рис. 4). Структура построенной онтологии соответствует структуре и классификации элементов материаловедческой базы знаний.

Основное окно программы Protege состоит из закладок, которые отображают различные аспекты модели знаний.

Классы (Classes)

Классы соответствуют объектам или типам объектов предметной области СТМ. В нашей онтологии классы включают в себя типы поликристаллов синтетических алмазов, отличающиеся по способу изготовления и исходным материалам. В подкласс исходных материалов включены порошки графита и порошки синтетических алмазов. Кроме того, способы и режимы получения ПКА также представлены в виде классов объектов.

Классы отображаются в виде иерархии наследования. Свойства классов, выбранных в текущий момент в навигаторе, отображены в редакторе классов (рис. 5).

Для решения задач многофакторного анализа и прогнозирования физико-механических свойств СТМ необходимо описать все технологические аспекты их получения, структурные особенности исходных порошков, а также физико-механические характеристики полученных материалов. Поэтому классы «Порошки синтетических алмазов с добавками» и «Порошок графита» имеют подклассы, содержащие объекты, отличающиеся зернистостью исходных порошков. В свою очередь, следующие подклассы содержат более конкретные объекты – экземпляры.

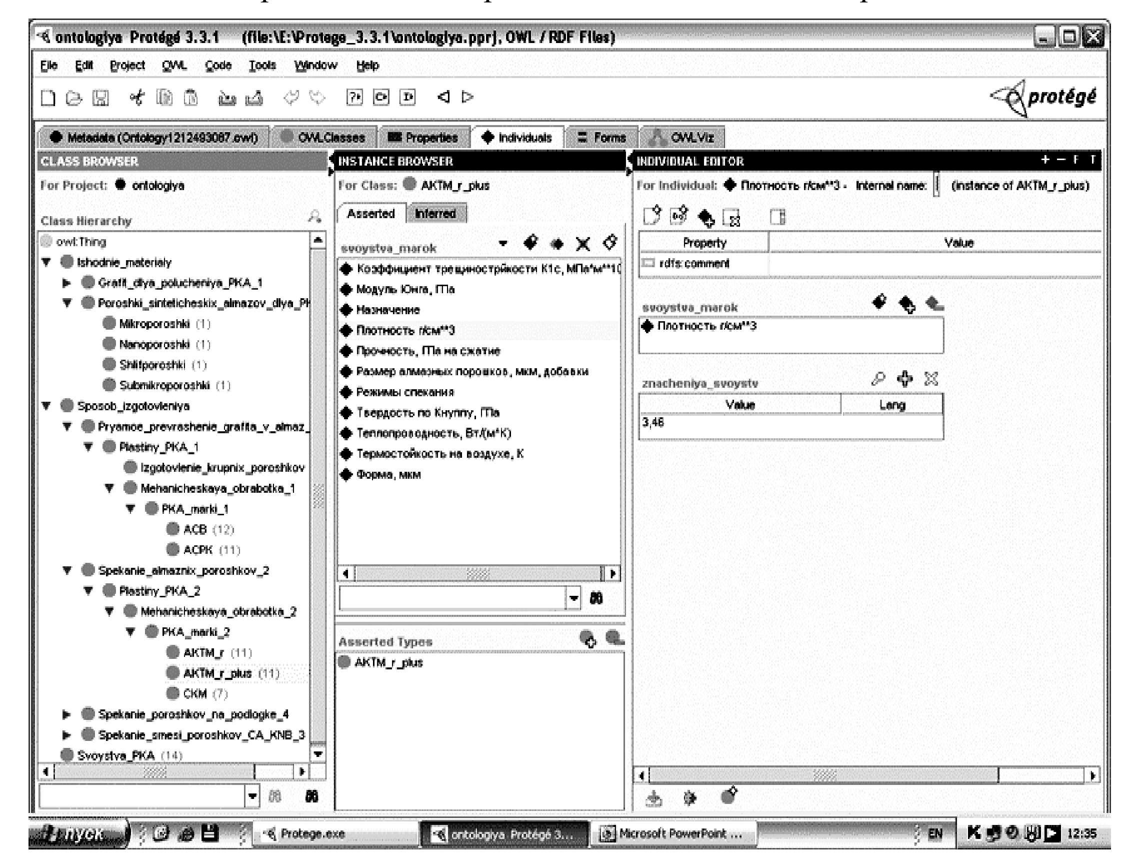


Рисунок 5 – Иерархия классов онтологии «Поликристаллы синтетических алмазов»

Создание экземпляров классов

Экземпляры классов – это данные базы знаний ПрО. Перед вводом конечных данных экспертами была проведена окончательная проверка структуры проекта и уточнена вводимая информация. Вид редактора класса с экземпляром показан на рис. 6. Например, подкласс «Микропорошки» содержит четыре экземпляра.

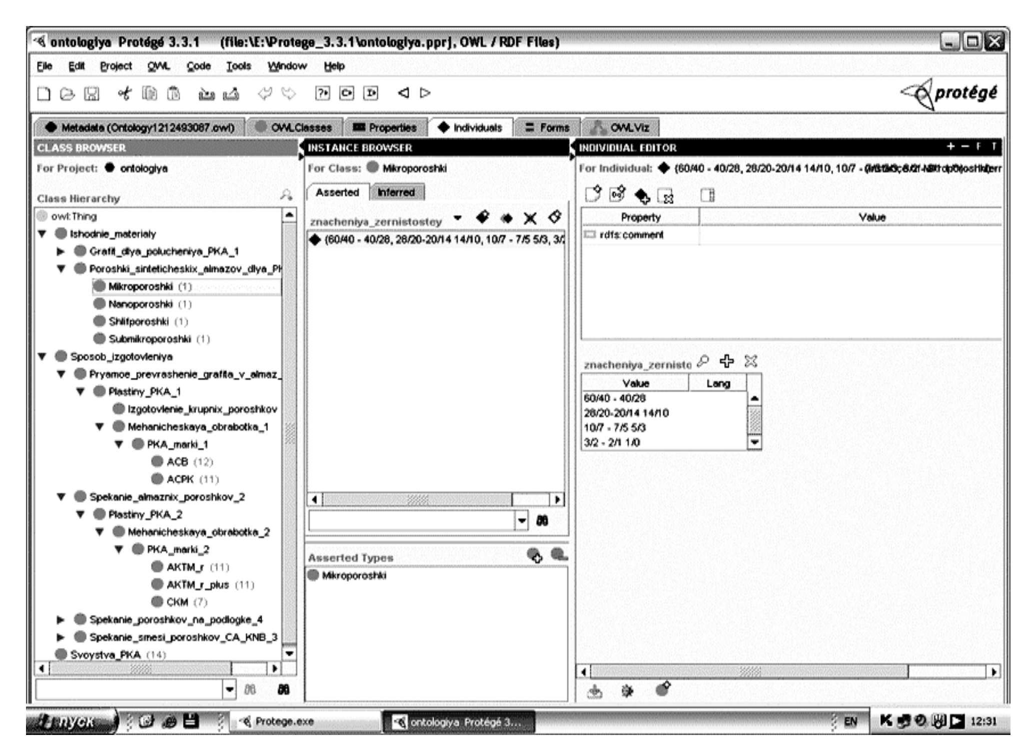


Рисунок 6 – Вид редактора подкласса «Микропорошки» с экземплярами

Создание слотов

В системе Protégé под классами понимаются конкретные понятия (концепции) предметной области, такие, как микропорошок или пластины ПКА.

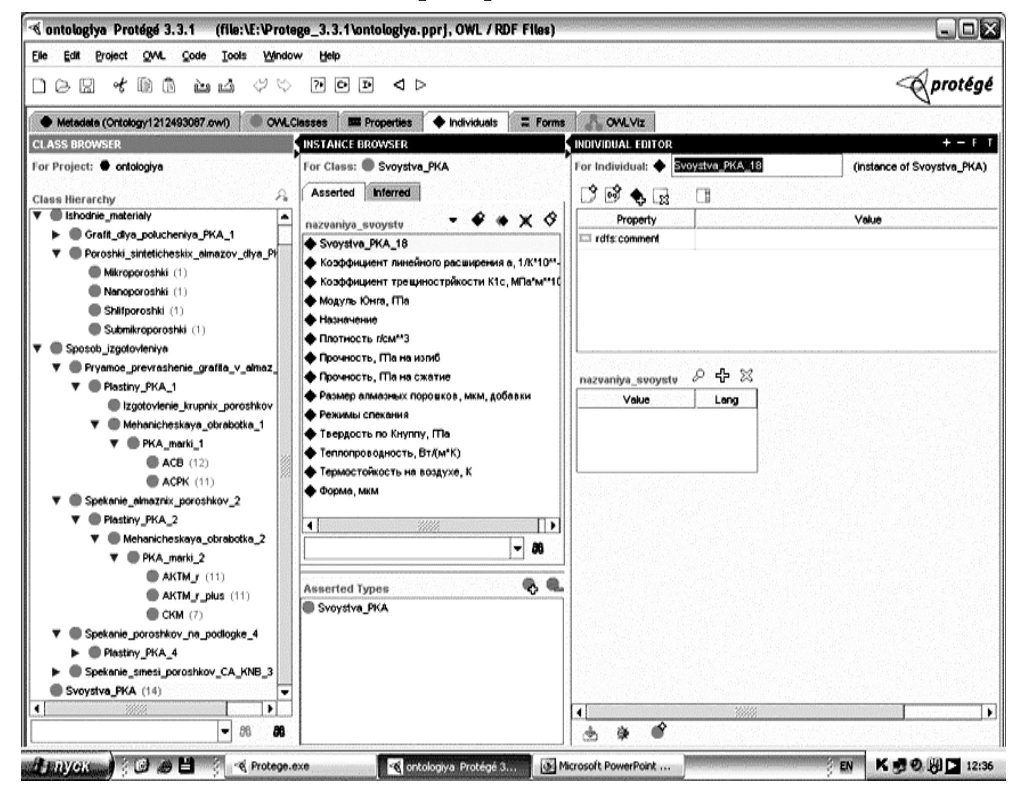


Рисунок 7 – Слот «Свойства ПКА»

В то же время классы – это больше, чем объекты, объединенные в иерархию. Они имеют атрибуты, в нашем случае – это физико-механические свойства. Эти атрибуты представлены в слотах, привязанных к классам и описывающим отношения между классами. Для создания слота «Свойства ПКА» были выбраны наиболее важные характеристики материалов (всего 14 характеристик).

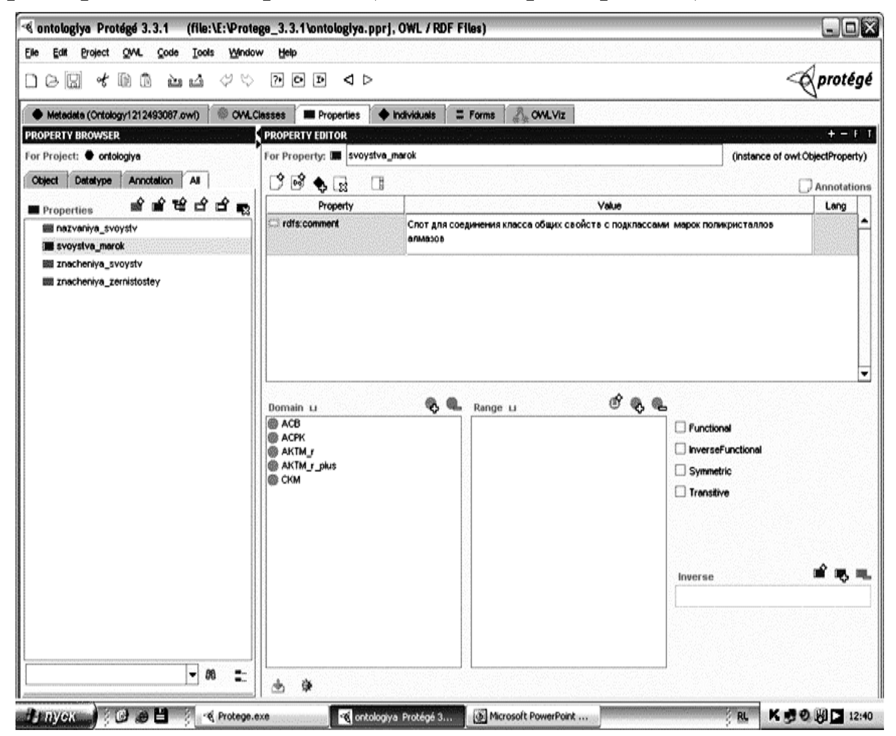


Рисунок 8 – Связывание объектов с их свойствами с помощью слота

Любой подкласс класса автоматически наследует все слоты базового класса.

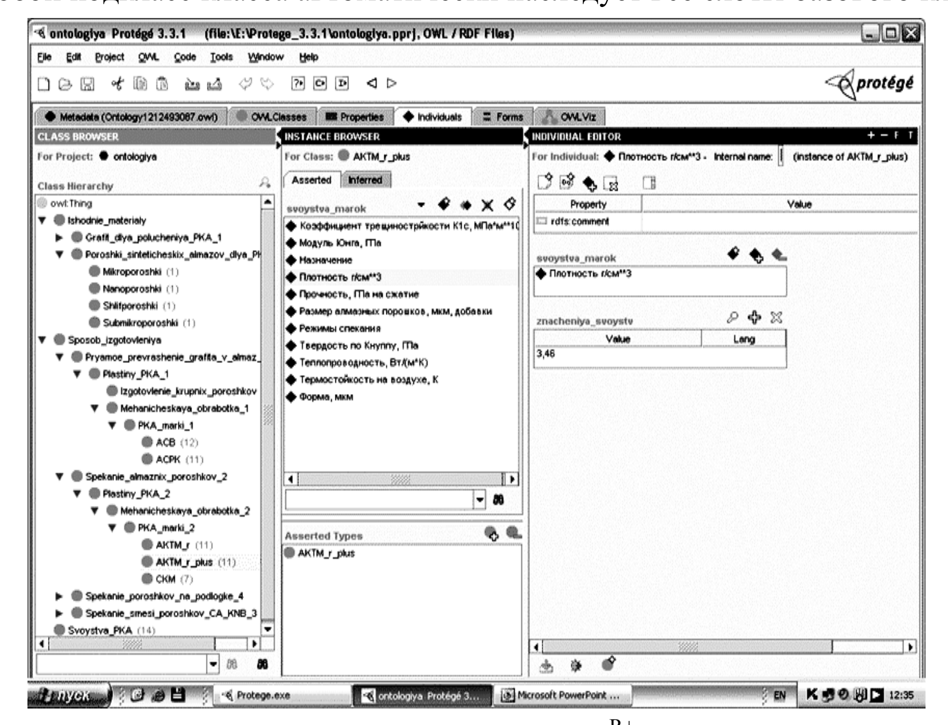


Рисунок 9 – Экземпляр АКТМ$^{R+}$ и его свойства

Подклассы более чем с одним базовым классом наследуют слоты от всех базовых классов. Множественное наследование – одна из основополагающих возможностей Protégé.

Система Protégé также позволяет создавать слоты, которые могут быть использованы для описания отношений между классами, которые не определены в иерархии классов. Для этого существуют слоты Instance (экземпляр) или Class.

Создание и сохранение запросов

Закладка запросов позволяет пользователю получать сведения из проекта по всем экземплярам классов, которые удовлетворяют интересующим критериям. Для этого создается запрос, в котором выбирают один или более классов и один или более слотов в классе. Запросы могут быть сохранены в библиотеке для последующего использования. На рис. 10 показана панель отображения результатов поиска по значению теплопроводности.

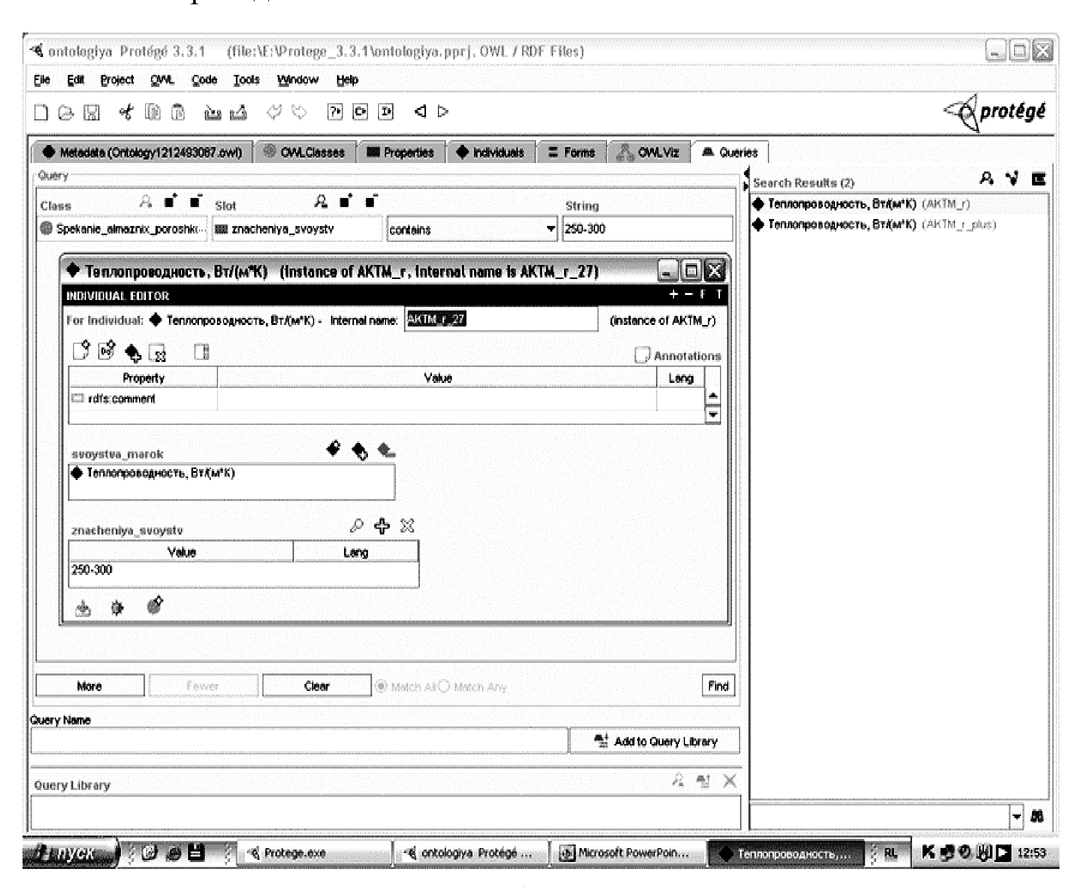


Рисунок 10 – Панель отображения результатов поиска

Таким образом, в результате проведенных работ разработана модель предметной области «Сверхтвердые материалы», включающая систему дополняющих друг друга многоуровневых онтологий. Система материаловедческой онтологии «Сверхтвердые материалы» содержит понятия проблемно-ориентированных знаний, соответствующих научным направлениям, разрабатываемым в институте сверхтвердых материалов. В процессе разработки данной модели использовались базы данных по прикладным задачам рассматриваемой предметной области. С использованием системы Protégé

спроектирована онтология «Сверхтвердые материалы». В статье дается описание процедуры проектирования небольшого фрагмента иерархической модели предметной области «Поликристаллы синтетических алмазов».

Литература

1. Артемьева И.Л., Рештаненко Н.В. Интеллектуальная система, основанная на многоуровневой онтологии химии // Программные продукты и системы. – 2008. – № 1. – Режим доступа: http://www.swsys.ru/index.php?page=article&id=113.
2. Попов Э.В. Корпоративные системы управления знаниями // Новости искусственного интеллекта. – 2001. – № 1. – С. 14-25.
3. Гладун А.Я., Рогушина Ю.В. // Корпоративные системы. – 2006. – № 1. – Режим доступа: http://www.enterprise.com.ua
4. Кулаковский В.Н., Фидаров Т.З. Структура и организация компьютерной информационно-аналитической системы для развития направления перспективных инструментальных материалов и инструментов из них в Украине // Мат-лы Междунар. конф. «Общество, основанное на знаниях: новые вызовы науке и ученым». – 2005. – Киев: «Феникс». – С. 360-368.
5. Лебедева А.А., Скворцов И.В., Фидаров Т.З. Формы представления знаний и алгоритмы принятия решений в информационных экспертных системах в производстве и применении сверхтвердых материалов и инструментов из них // Мат-лы Седьмой ежегодной Междунар. Промышленной конференции «Эффективность реализации научного ресурсного и промышленного потенциала в современных условиях». – Славское, Карпаты, 2007. – С. 125-128.
6. Лебедева А.А., Фидаров Т.З., Скворцов И.В. Подход к построению предметной онтологии для экспертной системы «Сверхтвердые материалы» // Сб. научных трудов «Породоразрушающий и металлообрабатывающий инструмент – техника и технология его изготовления и применения. – Институт сверхтвёрдых материалов им. В.Н. Бакуля НАН Украины, Киев, 2007. – Вып. 10. – С. 370-374.
7. Наталья Ф. Ной (Natalya F. Noy), Дэбора Л. МакГиннесс (Deborah L. McGuinness). Разработка онтологий 101: руководство по созданию Вашей первой онтологии. – Стэндфордский университет, Стэнфорд, Калифорния, США. – Режим доступа: http://protege.stanford.edu/publications/ontology_development/ontology101.html

В.Н. Кулаковський, А.А. Лебедєва, К.З. Гордашник, Є.М. Чистякова, І.В. Скворцов

Онтологічний підхід до побудови бази знань «Надтверді матеріали»

Запропоновано підхід до побудови багаторівневої онтології для системи знань предметної області «Надтверді матеріали (НТМ)». Розглядається складноструктурована предметна область надтвердих матеріалів, визначаються компоненти та процес розробки бази знань НТМ на основі метаонтології «Надтверді матеріали (НТМ)». Авторами розроблено модульний підхід до створення онтології та показано його застосування на прикладі невеликого фрагмента ієрархічної моделі предметної області «Полікристали синтетичних алмазів». Пропонується створення онтології у вигляді сукупності модулів, де кожний модуль описує термінологію деякого розділу предметної області. В роботі надається стислий аналіз предметної області «Надтверді матеріали», її ділення на розділи, описуються онтології та моделі двох розділів – «Природні НТМ» та «Синтетичні НТМ».

V.N. Kulakovskiy, A.A. Lebedeva, K.Z. Gordashnik, Ye.M. Chistikov, I.V. Skvortsov

The Ontological Approach to Constructiny of Knowledge Base «Superhard Materials»

The authors offer an approach to constructing a multilevel ontology for the system of knowledge of the «Superhard materials (SHM)» subject domain. The complex-structure subject domain of superhard materials is examined; components and the process of development of the knowledge base are defined on the basis of the «Superhard materials (SHM)» metaontology. The authors' modular approach to building an ontology model and its application are illustrated with a fragment of the hierarchical model of a domain named «Polycrystals of synthetic diamonds». The ontology is built as a set of modules, where each module describes terminology of a particular section of the subject domain. The authors briefly analyze the «Superhard materials» domain and describe the ontology and models of two domains – «Natural SHM» and «Synthetic SHM».

Статья поступила в редакцию 22.07.2008.